



King's Research Portal

Document Version
Peer reviewed version

[Link to publication record in King's Research Portal](#)

Citation for published version (APA):

Yeung, K. (2016). The Forms and Limits of Choice Architecture as a Tool of Government. *Law and Policy*.

Citing this paper

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

General rights

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



The Forms and Limits of Choice Architecture as a Tool of Government

Journal:	<i>Law & Policy</i>
Manuscript ID	Draft
Manuscript Type:	Original Manuscript
Keywords:	Coercion, inducements, nudge, liberty, tools of government, choice architecture, deception, autonomy

SCHOLARONE™
Manuscripts

Review Only

The Forms and Limits of Choice Architecture as a Tool of Government

by Karen Yeung

1. Introduction

Design has long been used to influence human behaviour. Yet “design-based” control techniques, broadly understood as the purposeful shaping of the environment and the things and beings within it towards particular ends (Yeung 2015) have, until the publication of Thaler and Sunstein’s highly readable paperback, *Nudge* (Thaler and Sunstein 2008) remained neglected as a focus of regulatory scholarship (cf Yeung 2008). *Nudge* popularized findings from laboratory experiments by cognitive psychologists which highlight considerable divergence between the rational actor model of decision-making assumed in microeconomic analysis and how individuals actually make decisions due to their pervasive use of cognitive shortcuts and other decision-making heuristics. The implications of these experiments can, as *Nudge* emphasizes, be readily harnessed by those with a stake in influencing the behavior of others through the use of “choice architecture” often in simple, inexpensive yet highly effective ways. Hence the mere rearrangement of items on a menu of options can systematically influence individual decision-making. For example, restaurant owners wishing to sell more of their most profitable dishes could list them at the beginning or end of the menu, given that items so placed are twice as popular compared with their popularity when the same items are placed at the centre of the menu (Dayan & Bar-Hillel 2011). To increase the average price of the bottles of wine sold, restaurant owners can add a few more expensive bottles to the wine list: because people take “cues” from the list as to how much they should expect to pay for a bottle of wine, many patrons will buy a wine priced just below the most expensive on the list because they do not wish to be extravagant whilst not wishing to be seen as “cheap” (Metrix 2014). But it is not private dining establishments whom Thaler and Sunstein primarily seek to address, although commercial firms could utilize nudge¹ and other forms of choice architecture to enhance sales. Rather, their exhortations are targeted primarily at public policy-makers, urging them to employ nudge techniques to improve peoples’ decisions about “health, wealth and happiness” (Thaler and Sunstein 2008:14-15).

* Professor of Law and Director, The Centre for Technology, Ethics, Law & Society (TELOS), The Dickson Poon School of Law, King’s College London and Honorary Senior Fellow, Melbourne Law School. I am grateful to John Coggan, Shanmugapriya Umachandran and two anonymous referees for comments on earlier drafts. All errors remain my own.

¹ Thaler and Sunstein define a nudge as ‘an aspect of choice architecture that alters people’s behaviour in a predictable way without forbidding any options or significantly changing their economic incentives (Thaler and Sunstein 2008, 6).

Underlying Thaler and Sunstein's proposals is a simple yet important insight: that choice architecture is ubiquitous and unavoidable. But does it follow, as Thaler and Sunstein suggest, that policy-makers can utilise choice architecture as they see fit? If choice architecture is as benign and potentially welfare-enhancing as a cursory read of *Nudge* might suggest, then one might allow the state a free reign in utilizing it to implement their policy goals. But there are other forms of choice architecture besides nudge. The gunman who offers his victim, "your money or your life?" is as much a choice architect as the cafeteria manager who places the fruit at eye level while placing the chocolate cake further back to encourage patrons to make healthier dietary choices and the supermarket owner who slashes grocery prices on their "use by" date to stimulate sales. This paper focuses on the following three forms of choice architecture that may be employed by the state in order to influence the behavior of others:

- *coercion*: a technique in which the choice architect (D) seeks to compel another (V) to perform a particular action by bringing about, or threatening to bring about, some unwelcome consequence if V does not perform that action;
- *inducements*: a technique through which the choice architect (D) seeks to encourage another (V) to engage in a desired action, by offering to provide some kind of benefit to V if she engages in that action; and
- *nudge*: a technique through which D seeks to encourage another (V) to engage in a desired action by intentionally arranging the choice environment to render it systematically more likely that V will take the action D desires whilst allowing V to easily and cheaply avoid taking the action preferred by D.

One of my central concerns is to evaluate whether these forms of choice architecture can be understood as "libertarian": by this I mean that they cohere with the fundamental values and premises upon which liberal democratic states rest (Waldron 1987). Chief among these values is the importance of individual liberty and freedom and the concomitant special status accorded to individual choice in liberal democratic communities (Mill 1859; Scanlon 1988). In so doing, I will highlight different ways in which these techniques may be regarded as an interference with individual freedom, and the conditions under which such interferences might be rendered acceptable or otherwise justified.

This paper proceeds in four parts. First, I begin by critically interrogating Thaler and Sunstein's claim that because choice architecture is ubiquitous and inescapable, the state can utilize choice architecture as it sees fit. Secondly, I analyse the underlying mechanisms of each architectural form through the lens of the liberal commitment to liberty and freedom in which interferences with individual choice are presumptively invalid. Thirdly, I explore the conditions upon which

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

these concerns might either be overcome or otherwise justified, focusing upon the state’s motive in seeking to prevent harm to others, and in protecting individuals from self-harm (i.e. paternalism). Finally, I conclude by briefly identifying how my insights might be further developed by drawing attention to shortcomings of a liberal analysis of choice architecture.

Before proceeding, the parameters of this inquiry should be noted. Although my analysis is general in nature, my primary concern is to identify whether the liberal democratic state can legitimately adopt such techniques to implement its public policy objectives. Hence I do not address the legitimacy of private sector use of choice architecture, although my critique may well have implications beyond the public policy context. Also excluded is the design and architecture of *public goods*, that is, goods for which consumption is both non-excludable (so that once the good is provided to a single consumer it is not possible to exclude consumption by others) and non-rivalrous (i.e. one person’s consumption of the good does not affect the ability of others to consume that good) (Sandmo 2008). Public goods include both tangible public infrastructure goods (such as public footpaths, streets, roads, motorways, national parks and communal recreational facilities) as well as largely intangible public goods (such as the presence and protection of the rule of law within a particular community, or the provision of independent assurances about the safety or quality of particular products available for consumption). Because the provision of public goods generates several important challenges which warrant extensive analysis (eg Arnold 2009:159), this paper focuses on *individual* choice architecture – that is, choice contexts that frame individual decision-making, whether offered on a standardised basis to all affected individuals (such as financial incentives offered to property owners who install solar energy panels on their roof-tops) or on a one-to-one basis in which choice architecture is tailored and directed at a particular individual (such as the medical treatment options offered to a specific patient).

2. Can the state design choice architecture as it sees fit?

Given that choice architecture is inevitable and unavoidable, Thaler and Sunstein argue that policy-makers have free reign to design choice architecture except in undefined circumstances in which “active choosing is sometimes the right route” (Thaler and Sunstein 2008: 243). This reasoning is flawed for at least three reasons. First, it overlooks the normative significance of the distinction between accidental and intentional action (Lamond 2000: 54). As Hansen and Jespersen point out, there is a fundamental distinction between a given context that *accidentally* influences behaviour in a predictable way and someone – a choice architect – *intentionally* trying to alter behaviour through the design of those contexts (Hansen and Jespersen 2013: 10). The distinction between intentional and accidental action is critical to normative evaluation,

1
2
3 profoundly affecting our conceptions of agency and responsibility and which runs right through
4 the analytical heart and structure of the criminal law. As Oliver Wendell Holmes Jr famously
5 stated, “even a dog knows the difference between being kicked and stumbled over” (Wendell
6 Holmes Jr 2004: 3). There is a world of difference between my response to the injury which I
7 suffer from tripping over a rock which is inadvertently left lying on the ground, and one which
8 has been deliberately placed in my path in order to trip me up. Although I am entitled to
9 demand an explanation and seek to identify those whom I regard as responsible for my injuries
10 in both cases, only in the latter case will the perpetrator’s conduct attract strong condemnation
11 as morally reprehensible (Cane 2002). This is equally true of state actors who use choice
12 architecture to implement policy goals.²

13
14
15
16
17
18
19 Secondly, the liberal democratic state has an obligation to explain and justify the nature and
20 character of social arrangements to those who are subject to them. Hence rather than
21 dispensing with the need for proper justification, the ubiquity and inescapability of choice
22 architecture makes it all the more important that policy-makers who intentionally utilise it in
23 furtherance of particular ends offer a principled justification for their actions. As Jeremy
24 Waldron observes, despite considerable differences between theories that proceed under the
25 banner of liberalism, liberals share a common commitment to a conception of freedom and of
26 respect for the capacities and the agencies of individual men and women, grounding a
27 requirement that all aspects of the social should either be made acceptable or be capable of
28 being made acceptable to every last individual (Waldron 1987). Hence liberalism rests on an
29 enduring commitment, often expressed in the idea of a social contract, that the social order
30 must be one that can be justified to those who live under it. A liberal society aspires to be a
31 transparent order in that its workings and principles should be well-known and available for
32 public apprehension and scrutiny (Waldron 1987: 146). Given the normative significance of the
33 distinction between intentional and unintentional action, I will confine my attention to
34 *intentional* choice architecture, adopting the definition of nudge proposed by Hausman and
35 Welch as an “intentional intervention enacted by one party systematically to influence the
36 choice of another party” rather than adopting Thaler and Sunstein’s view that that, because
37 contexts always influence individual behaviour in a predictable way, we are always being nudged
38 (Hausman and Welch 2010). Thirdly, nudge is only one form of choice architecture. As I
39 shall demonstrate, other architectural forms may interfere with individual choice in legally and
40
41
42
43
44
45
46
47
48
49
50
51
52

53
54 ² Hence I disagree with Quigley who poses the question – do we have reasons to prefer choice
55 architecture that results from countless random influences or that which has been deliberately designed,
56 and argues that (at least in relation to health), intentional choice architecture to be preferred: Quigley
57 2013. Her analysis overlooks entirely the normative significance of intentional vs accidental action,
58 particularly when employed by the state. In my view, only intentional choice architecture that can be
59 *adequately justified* to its citizens is to be preferred. See also Saghai 2013: 491.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

ethically significant ways. Hence it is to three different forms of choice architecture - coercion, inducements and nudge, to which I now turn.

3. Forms and limits of choice architecture

3.1 Regulators as choice architects

Julia Black defines regulation as the “sustained and focused attempt to alter the behaviour of others according to defined standards or purposes in order to address a collective issue or resolve a collective problem or attain an identified end or ends, usually through a combination of rules or norms and some means for their implementation and enforcement, which can be legal or non-legal” (Black 2008: 139). Central to this definition is intentional action aimed at changing the behaviour of others to attain some identified end. On Thaler and Sunstein’s understanding of a choice architect as one who “has the responsibility for organising the context in which people make decisions” (Thaler and Sunstein 2008: 3), the strategies adopted by choice architects to change the behaviour of others for particular purposes constitute *regulatory techniques*. This characterisation enables us to draw from the rich and growing literature concerned with critically evaluating their legitimacy (Morgan and Yeung 2007: chapter 6; Baldwin et al 2012: chapter 3; Lodge and Wegrich 2012: chapter 12) in order to identify whether, and to what extent, such techniques raise legal, democratic and moral concerns, emphasising the ways in which they can be understood as preserving or antagonising individual liberty and freedom. Although the three forms of choice architecture examined here are not intended to provide an exhaustive typology, they nonetheless cover a wide range of social choice contexts, enabling the salient similarities and differences between each form to be identified and compared.

3.1 Preliminary considerations

Three related considerations help to orient this analysis: first, acknowledging that the social conditions under which individuals make choices may fall well short of some idealised choice context without raising moral or legal concerns; secondly, noting limitations associated with “sliding scale” approaches which arrange policy interventions along a continuum in accordance with their intrusiveness and the magnitude of costs (of whatever kind) they impose on choosers (Sunstein 2013), and thirdly, distinguishing between individual liberty, freedom and autonomy.

3.1.1 Idealised choice contexts

When interrogating the legitimacy of forms of choice architecture, it is helpful to have some concept of the idealised choice context in which an individual's decisions can be regarded as fully voluntary, so that a person may be regarded as having assumed the risks of harm arising from her self-regarding decisions. Feinberg argues that the voluntariness of one's choices is a matter of degree. At one end of the spectrum are acts and choices that he calls *perfectly voluntary* (or "deliberately chosen" as Aristotle termed them). These are decisions and actions of normal adult humans in full control of their deliberative faculties, when fully informed of all relevant facts and contingencies, with their eyes wide open, in which there is calmness and deliberateness, no distracting or unsettling emotions, no neurotic compulsion, and no misunderstanding (Feinberg 1989: chapter 20). At the opposite end of the spectrum are choices that are not voluntary at all: *involuntary acts or choices*, such as the bodily movements of a person who lacks all intentional muscular control or is knocked down, or who through ignorance chooses something other than what one means to choose such as, to use Feinberg's vivid illustration, thinking arsenic powder is table salt and thus choosing it to sprinkle it on one's scrambled eggs (ibid). But neither the law nor morality require idealised conditions of choice for them to qualify as legitimate and freely chosen. People frequently make free, rational decisions in less than ideal circumstances in ways that are neither wrongful or ethically problematic. Choices are always made from among a limited choice set or options and, so long as we interact with one another, the actions of others will affect one's choices (Wertheimer 1988). As White points out, if I make a bid of \$100,000 on an houseboat you can now no longer purchase it for \$80,000, and when I take the last seat at the bar, you have to stand or go find somewhere else to drink, yet in neither case is the restriction of choice legally or morally questionable (White 2010). In other words, the mere fact that choice architecture reduces a person's options is not, in and of itself, necessarily problematic.

3.1.2 Classifying choice architecture according to the magnitude of costs

Secondly, much of the literature on policy instruments focuses on their coerciveness, often arranged along a sliding scale from least to most coercive according to some natural ordering of steadily increasing burdens of proof and persuasion associated with the expansion of the government's role in private activities (Linder and Peters 1989: 46; Nuffield Council on Bioethics 2007: Chapter 3). Sunstein adopts this approach in his discussion of paternalistic interventions³, suggesting that they are best understood in terms of a continuum in accordance with the magnitude of costs (of whatever kind) imposed on choosers by choice architects (Sunstein 2013: 1859). Although the degree of pressure imposed by the choice architect, and the associated costs associated with avoiding the choice architect's preferred outcome, are

³ See section 4 below.

salient and useful touchstones for comparison, there are other conceptual features of importance that have legal, democratic or moral significance, which the following discussion highlights.

3.1.3 Freedom, liberty and autonomy

One important set of conceptual distinctions which a “sliding scale” approach to the classification of policy instruments overlooks is the different ways in which individual liberty or freedom can be understood. Although liberal theorists often refer to interferences with individual liberty or freedom, these are not the same thing, although the terms are often used loosely and interchangeably in both policy and academic discussion. Liberty typically refers to the absence of rule-imposed duties. For example, because the law specifies maximum speed limits for motor vehicles, I am not at liberty to exceed this speed, even in the early hours of the morning when there are very few vehicles about and my Maserati can easily attain much higher speeds. In contrast, freedom refers to the practical possibility of undertaking specified actions, rather than the extent to which such action is constrained by duty-imposing rules. Hence I may be at liberty to drive at 100 m.p.h on the M25 motorway but I may not enjoy the freedom to do so if my car is stuck in a traffic jam. In other words, freedom refers to the de facto absence of effective constraints on actual or possible choices (Feinberg 1989: 630). One of the strengths of *Nudge* lies in highlighting how the intentional use of choice architecture may systematically influence individual decision-making while formally preserving individual freedom of choice and without impinging upon individual liberty. Hence the driver approaching a speed hump installed on a residential street retains both liberty and freedom to continue driving at speed, even though this will result in considerable discomfort and risks damaging her vehicle. Yet because much academic discussion has tended to focus on the state’s invocation of legal coercion to influence individual behaviour, it has focused on interferences with *liberty* rather than their effect on individual *freedom*.

In contrast, the distinction between respect for negative and positive liberty is well-established and important, although the content and contours of each of these ideas remains highly contested (Berlin 1998; Christman 2005). Negative liberty refers to the absence of constraint or intrusion by others, while positive liberty, or ‘autonomy’ refers to the capacity to be one’s own person, to live one’s own life according to reasons and motives that one takes to be one’s own and not the product of manipulative or distorting external forces (Raz 1986). What matters is not merely absence of constraint but the *quality of agency* that is of critical importance. Yet even those who regard individual autonomy, rather than liberty as absence of constraint, as the central concern of liberal theory, they nevertheless differ in the extent to which they regard

1
2
3 autonomy as worthy of special protection and hence immune from being overridden by other
4 values. While this is not the place to embark upon an extended analysis of the significance of
5 autonomy, I will touch upon some of these debates in considering the extent to which policy
6 makers can legitimately employ choice architecture in order to encourage individuals to make
7 better *self-regarding* decisions.⁴ For now, it is sufficient to recognise that because choice
8 architecture alters the circumstances of my choice, it is therefore a *potential* interference my
9 individual freedom and autonomy, but not – unless those restrictions on choice are legally
10 proscribed – with my liberty. With these considerations in mind, we now turn our attention to
11 specific forms of choice architecture, beginning with the most well-known: coercion.
12
13
14
15
16

17 18 3.2 Coercion 19

20
21 Coercion involves the application of sufficient pressure to bear on another to force that person
22 to do as the first person wills (Lamond 2000: 44). Although it attracted attention by classical
23 scholars from Aquinas through to Hobbes, Kant and Locke, western philosophers did not
24 engage in sustained analytical attention of coercion until the 1970s onwards, commencing with a
25 seminal paper by Nozick (Nozick 1969) who offered a list of necessary and sufficient conditions
26 for judging the truth of the claim that D coerces V:
27
28
29
30

- 31 1. D aims to keep V from choosing to perform action A;
- 32 2. D communicates that claim to V;
- 33 3. D's claim indicates that if V performs A, then D will bring about some consequence
34 that would make V's A-ing less desirable to V than V's not A-ing;
- 35 4. D's claim is credible to V;
- 36 5. Part of V's reason for not doing A is to lessen the likelihood that D will bring about the
37 consequences announced in 3.
38
39
40
41
42

43
44 Although we need not delve too deeply into the general legal and philosophical literature on
45 coercion, several features of Nozick's account are worth emphasising. In particular, coercion is
46 a *communicative* and *intentional* activity. Because the mechanism through which coercion is
47 intended to work depends upon appealing to V's rational decision-making capacities in
48 response to the unwelcome threat, it is often described as *rational coercion*. As Scott Anderson
49 explains, coercion is a "success" term: if the recipient of a proposal intended to coerce does not
50 subsequently alter her behaviour from the course it was on prior to receiving the proposal, then
51 the recipient was not coerced by it, and no coercion took place (2010: 4).
52
53
54
55
56
57

58 ⁴ See Section 4 below.
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

In contrast, techniques of *physical coercion* (or physical compulsion) work on the body of those against whom they are directed (V) and is irresistible by them, exemplified in the use of handcuffs and locked rooms to restrain a person’s freedom of movement (Lamond 2000:20). Not all scholars treat such physical restraints as instances of coercion, although both rational and physical forms of coercion are used to prevent an agent from taking certain actions, and are often used together (Anderson 2015). Unlike rational coercion, physical coercion need not have an explicitly communicative dimension. For example, airport terminals often force passengers to proceed through a retail complex to provide them with unavoidable shopping opportunities, even though passengers may be unaware that there is no alternative route though to the flight departure gates. In contrast, concrete bollards installed around the perimeter of pedestrian precincts serve both purposes: physically preventing motor vehicles from proceeding and as a clear visual signal to drivers that they ought not proceed.⁵

One of the intractable difficulties concerns how to distinguish coercive threats from legitimate offers, and identifying the relevant baseline against which such offers should be assessed (Anderson 2010: 5-6). Some scholars distinguish them by reference to expectations about what a person would “normally” expect to happen in the absence of an allegedly coercive offer in seeking to determine whether D’s offer makes V “worse off” than the baseline. Others favour a “moral” baseline for judging whether the offer makes V worse off. For example, Wertheimer persuasively argues that whether D is making an offer or a threat depends upon whether D is morally required to undertake the action which he is offering to undertake: if so, then the offer is a coercive proposal (i.e. a threat) which serves to bar or mitigate the ascription of responsibility for V’s choice made in response to the threat. In other words, if D has no *right* to make the proposal, then D’s proposal is to make V worse off than V would be relative to the moral baseline position. So for example, if D offers to rescue V from drowning in the sea in return for payment of £1000, whether this conditional offer constitutes a coercive threat depends upon whether V has a right to be rescued without such payment: if so, then the threat is coercive and V would be released from any obligation to pay D for undertaking the rescue. Hence a plea bargain offered by a prosecutor to an individual charged with of a criminal offence in the form of a sentence discount if she pleads guilty rather than risk trial on a charge that may carry a more severe sentence is *not* coercive, assuming that the accused has no right to demand that the prosecutor proceed with the more lenient charge (in other words, it assumes that that the more serious charge is an appropriate one in light of the available evidence)

⁵ The absence of any explicit communicative message associated with physically coercive forms of architecture does not exclude their association with an implied expressive dimension. For example, critics of situational crime prevention techniques claim that the use of physical architecture in ways intended to reduce opportunities for crime express an understanding of individuals as incapable of exercising moral self-restraint: see Duff and Marshall 2000.

(Wertheimer 1988). To avoid difficulties associated with identifying the appropriate baseline, other commentators focused on the powers, intentions and activities of the coercer. So for example, Lamond argues that rational compulsion is *prima facie* impermissible because it involves the intentional imposition of a disadvantage on the affected person if she does not comply with the demand (Lamond 2000). Similarly, Anderson argues that coercion is best understood as one agent employing power suited to determine, through enforceable constraints, what another will or (more usually) will not do, where the sense of enforceability is exemplified by the use of force, violence and threats to constrain, disable, harm or undermine the agent's ability to act (Anderson 2010:6).

We need not take a view on which analytical approach is to be preferred because it is widely agreed that that coercion clearly interferes with V's choice so as to render it defective. Although some coerced actions may be non-volitional, in the sense that V's will is "overborne" as a result of D's threat, Wertheimer points out that standard cases of coercion involve situations of "constrained volition": D credibly threatens V with some harm if V does not do A, and V does A because V rationally regards doing A as the most attractive alternative under the circumstances. V knows what she is doing and means to do it. V may feel pressured, of course, but the pressure does not affect V's cognitive or volitional capacities. V's choice is *involuntary*, rather than non-voluntary (Wertheimer 1993). As a consequence, coercion is typically thought to diminish the agent's freedom understood as individual autonomy, negating V's responsibility for the decision while, for most scholars, the coercive action by D is regarded as *prima facie* wrongful and thus stands in need of justification. Yet identifying the wrongfulness of coercion independently of the choice being coerced is far from clear (White 2010), prompting Anderson to observe that perhaps it is sufficient to note that it might be better to say that coercion is a "very potent means, prone to abuse, and something that deserves ethical scrutiny wherever it is used, rather than saying that it is intrinsically (even *prima facie* or *pro tanto*) immoral" (Anderson 2011).

3.3 Nudge

If forms of architectural choice are arranged on a sliding scale of intrusiveness (Sunstein 2013: 1859), then coercion and nudges would lie at opposite ends of the spectrum (Saghai 2013: 13). Unlike coercive offers, which involve the threat of unwelcome consequences of sufficient magnitude that any rational individual in V's position would seek to avoid, choice architecture aimed at intentionally influencing another's behaviour in the manner deemed desirable by the choice architect only qualifies as a nudge if V can easily and cheaply avoid D's preferred option. Thus it would not be irrational for V to choose to avoid D's preferred outcome. But a simple

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

comparison of these two forms of choice architecture by reference to the ease of avoiding the choice architect's preferred option overlooks the variety of nudge techniques and their moral and constitutional significance, including the ways in which some nudge techniques might fail to respect V's individual liberty. These concerns can be illuminated by drawing on insights from cognitive and behavioural psychology which are known as "dual process theory" and which lie at the intellectual foundations of Thaler and Sunstein's Nudge proposals (Kahneman 2013). This theory asserts that the human brain utilises two kinds of thinking: one, which is intuitive and automatic (which Kahneman calls "System 1" thinking), and another which is reflective and rational (which Kahneman calls "System 2" thinking). System 1 thinking is typically fast and instinctive, while System 2 thinking involves the deliberate, conscious processing of information and, relative to System 1 thinking, is slow, requires considerable effort and mental concentration. Any given behaviour can result from either mode of thinking: for example, one normally blinks automatically and without thinking, but one might also blink as a result of a conscious and intentional decision to do so (Hansen and Jespersen (2013): 13-14).

We can then combine insights from dual process theory with Hansen and Jespersen's distinction between *actions* and *causes*. Actions are often defined in terms of the "states-of-the-world" that an agent intentionally seeks to bring about: an intention resulting from the active deliberation about what the agent believes to be the available courses of action in the situation, and determined by her preferences over expected consequences associated with each option. The resulting bodily movement, inference or judgment that follows from this cognitive process can be referred to as an action, and is described as a consequence of a process that we call a *choice*. In other words, a choice may reasonably be interpreted as an end-result of the intervention of reflective (System 2) thinking. In contrast, "non-voluntary" or "involuntary" actions are those that are not considered as "real" actions, but only so in a derivative sense as being behaviour under the potential control of a given agent. Such "non-voluntary" actions can be understood as an event that happens to an agent, but which the agent could have controlled, such as blinking when a ball is thrown at you (reflexes), covering your mouth when you cough (habit) or a chess master immediately spotting the strong moves in a chess game (expertise) (Hansen and Jespersen 2013: 14). Because such non-voluntary "actions" are unintentional and do not involve active deliberation, they are usually not conceived of as resulting from choice as it is typically understood. Rather, Hansen and Jespersen claim that they are generally regarded as having been "caused" by other events, and – at least in the cases of reflex and socialised habits, the agent involved is typically not asked to assume responsibility for the "action" in question, which are strikingly similar to behaviours resulting from automated or System 1 thinking. On this view, Hansen and Jespersen suggest that concepts of "choice" and "action"

can be confined to those movements or processes which result from System 2 thinking. While some nudges seek to facilitate or build upon the kind of reflective thinking associated with System 2, many nudges are aimed at influencing the behaviour maintained by System 1 thinking.

Nudges that seek to utilise or facilitate System 2 thinking include those that seek to promote or build upon thoughtful, reflective deliberation. These include measures intended to help individuals comprehend the range of options available (such as government information campaigns intended to provide individuals with information that will help inform their decisions) or which seek to facilitate conditions for informed, thoughtful decision-making (such as laws requiring a minimum “cooling off period” following certain kinds of transactions that are believed to carry a significant risk of generating undue burdens, or which fail to reflect individual’s more considered choices).⁶ Nudges of this kind are liberty-promoting, helping ensure that individual decisions are made on an informed, reflective basis in a manner that more closely resembles the idealised choice context than might otherwise prevail. Similarly, nudges that take the form of physical architecture which seek to provide visible cues that prompt *conscious* changes in decision-making (such as the speed hump which operates by highlighting the costs and consequences of alternative courses of action to the approaching driver) do not “interfere” with individual liberty, although they clearly affect the individual’s decision-making options.

Yet, as Mols et al argue, Thaler and Sunstein’s conception of nudge encompasses both time-honoured techniques of social influence and persuasion which seek to appeal to people’s self-understanding as individuals who endeavour to “do the right thing”, and the intentional manipulation of the choice environment covertly to “trick” others into conforming with desired behavioural norms (Mols et al 2015). The latter kind of nudge typically utilise defaults and anchors in order to enlist System 1 thinking, rather than engage processes of reflection and evaluation. It is the behavioural tendency to “do nothing” that makes the default option “ubiquitous and powerful” (Thaler & Sunstein: 93) an effect which can be harnessed and magnified by policy-makers if combined with some implicit or explicit suggestion that it represents the normal or even the recommended course of action (ibid). Nudges designed into the physical architecture which seek to harness weaknesses or irrationalities in human cognition, such as the fly image etched into the base of the urinals at Amsterdam’s Schiphol airport which are intended to “improve the aim” (Thaler & Sunstein 2008: 4 and 91), constitute conceptual analogues. It is these kinds of nudges, which Mols and his colleagues term “governance by

⁶ It is questionable whether legislatively mandated “cooling off” periods constitute a nudge if they cannot be waived by the beneficiary.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

stealth” mechanisms (Mols et al 2015) have attracted the most extensive criticism, with scholars arguing that because they seek to influence individual decision making by exploiting unavoidable cognitive biases, they involve a form of invidious manipulation, deliberately seeking to by-pass the individual’s rational decision making processes. Such techniques are alleged to express contempt and disrespect for individuals as rational beings, failing to demonstrate respect for individual freedom understood as respect for individuals as autonomous, self-reflecting and rational beings capable of reasoned decision-making (Waldron 2014; Yeung 2012). Unlike coercion, in which V makes a rational informed, active choice to behave in the manner desired by the choice architect to avoid the threatened consequences of failure to comply, these kinds of nudges may render an individual’s apparent choices defective because they may be regarded as *non-voluntary*: such choices can be understood as the result of an event that happens to you, but which you could have controlled, such as blinking, such behaviours are largely akin to those resulting from automated cognitive processes associated with System 1 thinking. Unlike coerced decisions, which are *involuntary*, individual decisions taken under the influence of System 1 nudges may be defective because they are *non-voluntary*: as analogous to decisions made on the basis of intentional deception that typically mitigate the individual’s responsibility for the affected decision.

On this basis, the central objection to “stealth” nudges of this kind lies in their allegedly deceptive character.⁷ If so, then much of the illegitimacy associated with these kinds of techniques arguably lies in their lack of transparency, violating constitutional requirements that all governmental action should be transparent and open to public scrutiny thereby ensuring that the government is legally and democratically accountable for its actions. Can these concerns be overcome if the use of such techniques is publicly disclosed and rendered transparent and open to scrutiny? In my view, this depend largely on the *specific context* and purpose in which any given nudge is in contemplation or use, including whether the nudge infringes fundamental rights.⁸ Although there is vigorous contestation about the proper role, status and scope of rights in moral, political and legal discourse (Raz 1986, chapter 7; Dworkin 1985, chapter 3) fundamental (or human) rights can be understood here as delineating the minimum conditions under which human beings can flourish (that is, as moral agents) and which ought to be secured for them, if necessary by force (Kleinig 1978: 45-46). Thus, when confronted with social and political decisions which impact upon individuals, rights stand in the way of a simple consequentialist calculus, carrying special weight in a case of conflict with other social goals or

⁷ Although, as Wilkinson points out, the problem is in distinguishing manipulative and non-manipulative influences (Wilkinson 2013: 486).

⁸ See, for example, analogous debates concerning the morality of deliberate non-concealment of information by doctors with the intention of protecting patients from harm or distress (the so-called “therapeutic privilege”): Hodkinson 2013.

values and cannot lightly be overridden.⁹ Where no fundamental rights are at stake, full disclosure might be adequate to overcome the potential interference that such nudges may entail.¹⁰ Hence a highways authority could install a sign along a curved stretch of road stating something like the following:

Many drivers want to drive safely but often have difficulty judging the speed at which it is safe to drive along this stretch of road. In order to assist them, we have painted lines on the road in order to create the impression that the vehicle is speeding up, in order to help drivers reduce their speed to a safe level.

The problem is that while this kind of disclosure would overcome concerns about deception and manipulation, it may considerably reduce the effectiveness of the technique itself. As Bovens observes, many nudges “work best in the dark” (Bovens 2008). But in circumstances in which individuals have a right to be presented with choices in a particular form, typically grounded in a first-order fundamental right, a policy of disclosure will not overcome the deficiency. Consider for example, policies intended to lower the administrative costs and increase the efficiency of criminal proceedings: this might be achieved by advising individuals charged with a criminal offence that they are presumed to have waived their rights to a fair trial unless they expressly opt to put the prosecution to proof by returning a form to a court official within a specified time frame. Such a strategy is unlikely to be regarded as legally or constitutionally acceptable.¹¹ Although those charged with criminal offences might not be able to complain that they were not informed of their rights, such a regime could hardly be regarded as discharging the state’s obligations to guarantee their proper protection.¹² Of course, identifying whether an individual has a right to be presented with options in a particular form, grounded in a first-order right against the state to ensure its protection, will often be a contested matter. Nevertheless, a rights-based analysis provides a helpful framework for assessing whether

⁹ This special status is explored more fully in section 4.1.1(a) below.

¹⁰ While fundamental rights are justified by reference to deep and abstract notions concerning the inherent dignity of persons, the particular rights recognized by a society will invariably be a product of local, historical and cultural factors: Bleckman and Bothe (1986); Steiner (2000).

¹¹ Thaler and Sunstein demonstrate some awareness of rights-based constraints on the use of nudges in acknowledging that ballot papers should not be designed in favour of a particular candidate running for political office because ‘sometimes people have a right, even a constitutional right, to government neutrality of a certain kind’ (Thaler and Sunstein 2008: 246), and in their recognition that the constitutional right to freedom of expression provides the government from encouraging individuals to join a “Pray to Jesus Tomorrow” plan, or a “Dissent Less Tomorrow” Plan (ibid). Yet Thaler and Sunstein make no serious attempt to identify when such rights-based constraints on the use of nudge, and by extension, other forms of choice architecture, arise.

¹² Similar arguments were recognized by the UK Organ Donation Taskforce in considering whether to adopt a system of presumed consent to organ donation, concluding that it may entail a violation of the right to bodily integrity because it would result in the taking of organs from the disorganized: Organ Donation Taskforce 2008: 7 and 341.

objections to the use of nudge techniques that utilise System 1 thinking might be overcome or otherwise justified.

3.4 Inducements

If located on a sliding scale of cost and intrusiveness, (Sunstein 2013: 1859) in which coercion and nudges sit at opposite ends, choice architecture in the form of *inducements* would lie half way along¹³. Inducements are defined here as a technique through which the choice architect (D) seeks to encourage another (V) to engage in a desired action by offering V some kind of benefit if she engages in that action. The relevant benefit may take a positive form (such as a cash payment or other discount on purchase of some good or service that V desires) or a negative form (such as an offer to refrain from action that D could otherwise undertake, such as imposing an additional financial charge for failing to act in the manner D wishes to encourage, or refusing to grant some kind of benefit to V which D is empowered to withhold).¹⁴ Although choice architecture in the form of coercion and inducements both involve altering the costs associated with alternative options, in the case of coercion the costs imposed upon the V are regarded as so high that no rational individual in V's position could be expected to resist them (such as the gunman's offer of "your money or your life"), unlike inducements which might rationally be resisted despite the costs of resistance being considerable. Inducements can be distinguished from nudge techniques in at least two respects. First, an individual faces considerable costs in resisting inducements, unlike nudges, which allow V to avoid the nudge architect's preferred option easily and at relatively little or no cost. Secondly, while "stealth" nudges seek to harness instinctive System 1 cognitive processes, inducements seek to engage directly with an individual's capacity for rational reflective decision-making, thereby targeting System 2 cognitive processes.

Inducements are widely used as regulatory policy instruments, typically as financial incentives in the form of subsidies or tax benefits offered to those who undertake the desired behaviours (such as tax rebates for homeowners who install solar panels on their rooftops to generate electricity) or through the imposition of financial disincentives on those who engage in activities that the policy-maker seeks to discourage (such as the increasing size of road tax on the

¹³ These techniques are more often referred to in the regulatory and public policy literature as forms of "economic incentives" or "economic instruments" rather than inducements: eg Breyer 1982; Ogus 1994, chapter 11.

¹⁴ The distinction between negative and positive inducements is *not* coterminous with the distinction between a penalty and a tax: a tax is a financial imposed arising from legally permitted conduct, a penalty is a financial impost imposed for unlawful conduct: see Cooter 1984.

purchase of less fuel efficient motor vehicles). Such measures have attracted considerable scholarly debate, primarily from economists concerned to evaluate whether they produce economically efficient policies and outcomes.¹⁵ They have also been criticized on the basis that they have unfair, regressive and even discriminatory distributional effects, particularly in the form of so-called “sin taxes,” which involves levying charges on self-regarding actions that are considered undesirable (such as cigarette taxes) because they tend to impose disproportionate burdens on the poor.¹⁶ In a quite different vein, ethicist and philosopher Michael Sandel has cautioned against the use of financial incentives as an instrument for influencing the choices and behavior of others on the basis that it may denigrate their value (the “commodification” critique), inappropriately perverting or undermining the nature of certain goods and relationships and/or crowd out altruistic motives, intrinsic motivations and social solidarity (Sandel 2012; Grant 2011).

While these are all important considerations, my primary concern here is with interrogating whether using choice architecture as a regulatory policy instrument is consistent with demonstrating respect for liberty and freedom of choice. In this respect, inducements have been criticized as constituting a form of “undue influence” (Largent et al 2013). To evaluate this claim, it is helpful to compare inducements with coercive threats. They are similar in two respects. First, both are structured in the form of offers that can be expected to result in V feeling substantial pressure to accept the offer. But in the case of inducements, D has no legal or moral obligation to refrain from making the offer: hence D is legally and morally permitted to offer V a substantial discount on the price of an item if V offers to settle the debt immediately by tendering payment in cash, even though D is under no obligation to make, or to refrain from making, any such offers to V or to any other customer. In contrast, the gunman who threatens to shoot the bystander who refuses to hand over her purse is legally and morally obliged to respect V’s right to possession of her purse and its contents¹⁷. Secondly, in

¹⁵ When one party engages in an activity which imposes costs on another for which the latter is not compensated, the resulting “externality” may be corrected by imposing a tax on the party engaging in the harm-causing activity – these taxes are known as “Pigouvian” taxes, named after economist A.C. Pigou who observed that if such activities are unregulated, this is likely to result in excessive amounts of the activity in question (Pigou 1932: 172-203; Galle 2012).

¹⁶ Fairness in taxation policy is widely understood to involve equal treatment of persons in similar situations (horizontal equity) and differentiation in the burdens imposed on persons with varying abilities to pay (vertical equity). It seems intuitively unfair to impose the greatest economic burden on the poorest members of society, yet this is what sin taxes do because lower income users are “economically compelled” to desist, while those who can afford to absorb the extra cost suffer no real restriction of their range of consumption choices. See Haille 2009-2010).

¹⁷ See section 3.2 above.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

evaluating the legitimacy of both kinds of offers, it is important to distinguish the source of the pressure arising from the offer itself, and pressure arising from the background circumstances. Consider two individuals: X, who is a single parent who has lost her job, and Y who is suffering from a degenerative disease and wishes to have access to experimental drugs which have not yet received regulatory approval and which are therefore not yet available on prescription. Both are invited to participate in clinical trials on human subjects for which they will receive a modest cash payment. While it may be true that in both cases each may have no reasonable alternative but to accept the inducement and participate in the trials, the source of the pressure is not the offer itself, but the absence of reasonable alternatives (Largent et al 2013: 504). There are, as Largent and her colleagues point out, many situations in which people choose options because they lack reasonable alternatives without being coerced: a patient who agrees to surgery or chemotherapy because the only alternative is death is not coerced to consent and nor is her consent to treatment is invalid. Nor are those who take an unpleasant job to provide for their families coerced into doing so. Although such circumstances can be unfortunate and morally troubling, it does not necessarily make the medical care or employment unfair, and it certainly does not make them coercive.¹⁸

Unlike coercion, we lack a strong intuitive sense of what constitutes “undue influence”. In legal terms, undue influence is a doctrine developed by the courts of equity that applies wherever improper pressure (falling short of duress at common law) is brought to bear on a party to enter a contract or make a gift so as to render the contract or gift legally voidable.¹⁹ The influence may be “undue” by virtue of the actual circumstances and pressure brought to bear by one party over another, or by virtue of some kind of special relationship which exists between the parties in question.²⁰ The significance of the legal concept of undue influence in allowing the contract or gift made under undue influence to be avoided is a reflection of the law’s underlying concern that such transactions are made in the absence of a true or proper understanding of the nature of the transaction involved. In other words, undue influence is problematic because it distorts the individual’s capacity properly to comprehend the nature and magnitude of the consequences associated with the action or behaviour that is subject to influence, compromising the voluntariness of consent and thereby negating the individual’s decision-making responsibility (Largent et al 2013: 50). Where that influence takes the form of an inducement offered by D influence V’s conduct in the direction D prefers, such influence may be

¹⁸ Concerns about the misuse of power and the potential exploitation of the powerless associated with the use of incentives are thoughtfully analysed by Grant (2011).

¹⁹ *Bank of Credit and Commerce International v Abouody* [1990] 1 QB 923.

²⁰ *Barclays Bank Plc v O’Brien* [1994] 1 AC 180 and *Royal Bank of Scotland Plc v Etridge (No.2)* [2002] 2 AC 773.

considered undue if the inducement is so attractive that it distorts the V's evaluation of the risks and benefits of the relevant conduct (ibid: 505). Unlike coercion, which compromises the voluntariness of a decision (but typically not its rationality), undue influence can be usefully understood as compromising the *cognitive dimension* that is associated with rational, thoughtful deliberation (i.e. System 2 thinking). Individuals who are unduly influenced may act voluntarily, but their reasoning is distorted. For example, if the inducement offered to X and Y is such that it impairs, or could reasonably be expected to impair, their ability to exercise proper judgment such that they misunderstand the risks and benefits associated with participation in clinical trials either in terms of their magnitude, character and consequences such that they wrongly think that the benefit of payment exceeds the risks of participation, then the inducement may be regarded as interfering with their positive freedom (i.e. freedom as autonomy).

3.5 Summary

We have seen how three forms of choice architecture, coercion, inducements and nudge, may fail to demonstrate respect for individual freedom, but in different and quite distinctive ways. The individual who chooses to act in accordance with the coercive choice architect's desires to avoid the unwelcome threatened consequences does so on a rational, reflective and informed basis but which renders her decision defective and therefore involuntary. Although the individual whose decision-making outcome has been shaped through the use of a stealth nudge may act in a non-voluntary manner, the source of this lack of voluntariness lies in a form of deception that arguably fails to demonstrate respect for persons. Such deception need not be of concern if the choice architect's motives and techniques are properly disclosed to those encountering the nudge, except in circumstances where such techniques may involve an infringement of an underlying fundamental right. Inducements may render an individual's choice defective in a related but distinct sense: if the inducement is so attractive as to distort the individual's evaluation of the risks and benefits of the conduct in question, then the individual's decision to undertake the action to which the inducement is attached might be impugned on the basis that it is non-voluntary in a different sense. Even though it may have been intended by the choice architect to influence the chooser's rational decision-making faculties, it might nonetheless have served to distort her cognitive processes and thereby undermine the voluntariness of her choice.

4. Justifying choice architecture as a tool of government

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Although choice architecture may, depending upon its form, interfere with individual liberty and freedom, this does not *necessarily* imply that it is illegitimate. It merely demonstrates that it is problematic and prone to abuse. Accordingly, when employed as regulatory policy instruments, choice architecture must be justified and subject to institutional safeguards. What must be shown is that these techniques can be justified as a means for deliberately seeking to influence an individual’s decision-making in light of the motives of the choice architect, the nature and importance of the behaviour which the policy-maker seeks to encourage, the extent to which the technique undermines liberal democratic principles (including any interference with fundamental rights and the extent to which they distort or undermine individual freedom.

4.1 The importance of motive

In discharging this burden of justification, the state’s motives in seeking to influence individual behaviour is critically important. But the reasons or motives underpinning a particular policy measure are not always easy to discern, and policy-makers often seek to pursue multiple motives through a single policy instrument (Husak 2005; Pope 2004). What matters are the *reasons* supporting the introduction of the proposed measure, seen from the perspective of the policy-maker.²¹ Where a policy measure is plausibly underpinned by multiple motives, only one acceptable basis is required, it need not be justified on all possible grounds.²² Two broad policy motives are of particular salience here: the prevention of harm to others and the prevention of harm to self.²³

4.1.1 The Harm Principle

²¹ In order to overcome the practical difficulties associated with identifying the motives of legislatures and policy-makers in relation to a particular measure, Husak suggests that one could adopt an approach that sought to identify the “most plausible rationale” for the measure in question (Husak 2004). Similarly, Feinberg suggests that in situations where the legislature’s motives are difficult to discern, we can sometimes construct an ‘implicit rationale’ that provides it with a plausibly coherent rational reconstruction, the job that the law is tacitly understood to be doing (Feinberg 1989: 17).

²² Feinberg suggests that in cases where criminal prohibitions can be defended on two distinct grounds, both the need to protect individuals from the harmful consequences of their own acts *and* the need to prevent social harm generally, where the public interest is so clearly at stake that the paternalistic rationale is quite redundant and can be defended entirely on liberal grounds. But in modern western societies, it is presupposed that there is no necessity that public harm be caused in sufficient degree to implicate the harm principle whenever an individual deliberately injures himself or assumes a high risk of so doing (Feinberg 1989: Chapter 17, 22). Some scholars do not consider cases of “mixed paternalism” to be paternalistic at all, since they can be justified on other grounds (Nys 2008).

²³ Note that mechanisms for the provision of public goods are outside the scope of this paper, see section 1 above.

The most well-known and widely accepted grounds justifying state interference with individual liberty is the harm principle propounded by John Stuart Mill (Mill 1859). For Mill, “..the only purpose for which power can be rightfully exercised over any member of a civilized community, against his will, is to prevent harm to others.” Mill expressed this principle primarily in terms of the use of compulsion or coercion by the state “in the form of legal penalties or moral coercion of public opinion” (Mill 1859: 13) although his thesis was more generally concerned with the exercise of the state’s power over individuals (Mill 1859: 15). On this view, coercion by the state in either rational or physical form can be justified if necessary to prevent harm to others. For example, it is the harm principle that justifies most liberal democratic states’ public health legislation authorizing the forced detention and physical isolation of those with serious, highly contagious diseases (Ashworth and Zedner 2014: chapter 9). Similarly, nudges and inducements might also be justified by reference to the harm principle, at least on a “sliding scale” view of choice architecture. On this account, because coercion can be justified on the basis of the harm principle, and nudges and inducements impose less pressure on individuals to comply with the choice architect’s preferred behavior, these measures also fall within the scope of the harm principle.²⁴

(a) Justifying interventions with fundamental rights

In the preceding section, we noted that the intentional use of different forms of choice architecture might interfere with fundamental rights. Except in relation to a small class of absolute rights which admit of no qualifications (such as the right to freedom from slavery), a considerably more demanding form of justification is required to legitimize interference with fundamental rights.²⁵ The state must demonstrate that the intervention is necessary to secure a legitimate state objective, and that the scope and content of the interference is not a disproportionate means to secure that aim²⁶. For example, the European Court of Human Rights has held that injunctions issued by the Irish Supreme Court against two agencies which

²⁴ It is important to acknowledge that the scope, content and contours harm principle is itself indeterminate (Brownsword 2005).

²⁵ There is often considerable legal debate about the scope and parameters of the right and thus whether a given intervention infringes the relevant right. For example, the UK House of Lords held in *Austin v Commissioner of Police of the Metropolis* [2009] 1 AC 564 that the police practice of “kettling” public protesters by imposing a cordon on an area enclosing thousands of people who could only leave with the permission of the police was held to constitute a “restriction” rather than a “deprivation” of liberty and hence the measure did not engage the protection of Article 5(1) of the ECHR which provides that “[e]veryone has the right to liberty of security of person”.

²⁶ The European Court of Human Rights has interpreted the proportionality requirement such that the state must show that the measure was the least intrusive means for securing its legitimate aim: *Sunday Times v UK* [1979] 2 EHRR 245.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

provided abortion information and advice to women (but did not encourage or advocate abortion) forbidding them from telling Irish women 'the identity and location' of abortion clinics in Britain, violated the women's right to receive information, contrary to Art 10(1) of the ECHR and was not 'necessary in a democratic society for the protection of health or morals' in order to qualify as permissible under Art 10(2). In the Court's view, even if Ireland had a legitimate interest in protecting the life of the unborn (noting the Irish Constitution then guaranteed the equal rights to life of the unborn foetus and the mother) the injunction was disproportionate in its impact, because it prohibited counselling regardless of the age, health, or circumstances of pregnant women and posed a health risk to women, who would likely terminate pregnancies at later stages without adequate counselling.²⁷

Although the measure in question (i.e. the Irish Court's injunction prohibiting the provision of information concerning abortion services to women) is a form of legal coercion, it can also be understood as state-backed form of choice architecture intentionally intended to influence the decisions of individuals (primarily pregnant women) to refrain from seeking an abortion by shaping their informational environment, thus making it more difficult and costly to acquire accurate information about the availability of abortion procedures. It might therefore be regarded as a "nudge" or "inducement" depending upon the ease and cost which pregnant Irish women then faced in seeking to obtain reliable and accurate information about abortions in Britain and elsewhere. Because the justification offered by the state was alleged to lie in the protection of the unborn foetus, its motives for interference were claimed to rest primarily on the basis of the harm principle and to ensure due protection for the constitutionally recognised rights of the unborn, rather than as a paternalistic measure aimed at influencing the self-regarding decisions of pregnant women contemplating abortion.

(b) Legitimizing manipulative forms of choice architecture?

But what of "stealth" nudges, which are often claimed to involve a kind of deception that fails to demonstrate respect for individuals as rational, self-directing agents? Can these be justified on the basis of the harm principle? On the one hand, some might argued that the prevention of harm to others does not justify the use of deceptive techniques of any kind, because it fails to overcome the risk of abuse that such practices entail, particularly in the hands of the state. In particular, because these techniques are opaque and difficult if not impossible for citizens to detect, some claim we should rule them out altogether, lest we proceed down a slippery slope in which the state employs increasingly intrusive and manipulative forms of deception without

²⁷ *Open Door and Dublin Well Woman v Ireland* [1992] ECHR 68 (29 October 1992).

institutional means for systematically scrutinising their use, effectively conferring upon the state an unconditional licence to engage in deceptive practices to further its own ends (White 2013). On the other hand, if the kind of deception entailed by stealth nudges is relatively minor and is accompanied by proper safeguards, their use might be legitimate if employed to prevent significant harm. On this view, some kind of consequentialist balancing may be appropriate, in light of the wrong associated with the deception and its associated expression of contempt for individuals as rational agents, weighted against the benefits arising from nudging in the specific case, and the nature, character and severity associated with mistakes that could arise when defaults run contrary to any individual's preferred choice and that individual fails to opt out of the nudge's default. Consider the painting of lines on curved road bends to create the illusion that the vehicle's speed is increasing which involves a stealth nudge to prompt drivers to reduce their speed. If I slow down when driving because I fall prey to the optical illusion thereby created, the consequences of my mistake are very minor (I drive more slowly than I would otherwise have), as is the associated seriousness of the insult associated with the underlying deception, while the benefit is considerable: significantly reducing the risk of serious harm to myself and other road users. In these circumstances, stealth nudges might be justified, particularly if there is adequate transparency about their intended use enabling them to be critically scrutinised and challenged.

4.1.2 Paternalistic choice architecture

In defending the harm principle as the sole basis upon which the exercise of coercive state power over individuals can be justified, Mill vigorously attacked state intervention with an individual's self-regarding decisions. In an oft-cited passage in which he proclaims that "the only purpose for which power can be rightfully exercised over any member of a civilized community, against his will, is to prevent harm to others," he continues,

"His own good, either physical or moral, is not a sufficient warrant. He cannot rightfully be compelled to do or forbear because it will be better for him to do so, because it will make him happier, because, in the opinions of others, to do so would be wise, or even right. These are good reasons for remonstrating with him, or reasoning with him, or persuading him, or entreating him, but not for compelling him, or visiting him with any evil in case he do otherwise." (Mill 1859: 13).

The term "paternalism" did not appear until some time later, and it was not until the 1960s that lively scholarly debate concerning paternalism emerged in the legal and philosophical literature (Pope 2004). Although the exact boundaries of paternalism is a contested issue, Gerald Dworkin suggests that it involves at least three elements: (1) some kind of limitation on the freedom or autonomy of some agent; (2) without the consent of the agent; and (3) for a

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

particular class of reasons, that is, for the purposes of advancing the agent’s own good (Dworkin 2014).

(a) Liberal objections to paternalism

The liberal objection to paternalism is grounded in the importance of individual liberty and freedom. According to Mill, the fully voluntary choice or consent of a mature and rational human concerning matters that affect only his interests is such a precious thing that no one else (and certainly not the state) has a right to interfere with it simply for that individual’s own good (Mill 1859). In liberal democracies my choices are regarded as imbued with moral and political significance because they are *mine*: not primarily due to other qualities which might be judged morally desirable, such as their consequences in enhancing welfare, contributing to the public good, or because they display intelligence or good judgment. In societies which cherish individual freedom, it is presumed that an individual’s choices are worthy of respect, however foolish or unwise, and the liberal state cannot legitimately interfere with them without adequate justification and only if certain conditions are present (Raz 1986). The core liberal idea of personality, articulated in terms of personal autonomy, demands that individuals be allowed to choose and pursue their different plans or paths of life for themselves without interference from others (Kleinig 1983). Paternalism is problematic because it denies persons their individuality or standing as choosers, because it is generally misplaced since the paternalist largely lacks the subject’s concern for and knowledge of her own good, and because it neglects the developmental value of choice independent of the content and “correctness” of what is chosen (Kleinig 1983). Although respect for individual liberty and freedom is a fundamental cornerstone of liberal thought, there is considerable disagreement about the kinds of measures that are properly regarded as paternalistic, and whether and to what extent, the state may be justified in interfering with an individual’s self-regarding decisions. I will draw selectively from these debates to examine when and whether paternalistic forms of choice architecture can be considered legitimate.

On Gerald Dworkin’s definition, a policy measure only qualifies as paternalistic if it entails some kind of limitation on the freedom or autonomy of some agent to which the agent has not consented. While coercive choice architecture clearly constitutes such an interference, inducements that do not pose a significant risk of distortion to an individual’s evaluation of the risks and benefits associated with the activity which the choice architect seeks to encourage do not. Accordingly, ordinary democratic processes for scrutinizing governmental policy will suffice to ensure their legitimation – they need not satisfy the more searching demands of justification that apply to coercive or other policy measures that interfere with individual

freedom. Similarly, as we saw in the preceding section, stealth nudges are not as liberty-respecting as Thaler and Sunstein suggest because they involve a kind of deception, and/or because they entail an infringement of fundamental rights. But if these features are not present in a particular nudge technique, or if concerns about deception can be overcome through proper disclosure, then their use would not interfere with individual autonomy. Accordingly, the demanding form of justification needed to establish the legitimacy of interferences with liberty would not apply, so that ordinary democratic processes for scrutinizing public policy measures should also suffice (Harlow & Rawlings 2009; Parker 2016).

(b) Is the use of paternalistic choice architecture justified?

What of the states' intentional use of choice architecture to influence self-regarding decisions in ways that interfere with individual liberty or freedom? In what circumstances, if any, can state paternalism be justified within liberal democratic societies? Because this question has spawned lengthy and heated scholarly debate, I offer a more limited analysis here, drawing selectively from the work of legal philosophers Joel Feinberg and Gerald Dworkin to sketch two contrasting scholarly approaches, and also outline a middle-ground approach, reflected in the views of Andrew Simester and Andrew von Hirsh (Feinberg 1989; Dworkin 1988; Simester and von Hirsh 2011).

i. Feinberg's sovereign right conception of autonomy

Feinberg's treatment of paternalism in volume 3 of his magisterial work, *The Moral Limits of the Criminal Law*, arguably provides the most comprehensive contemporary analysis of paternalism. Feinberg adopts an explicitly "anti-paternalist" position, regarding all paternalistic reasons for intervention as invalid "by their very nature". This is because such interventions conflict with a "sovereign right" conception of personal autonomy in which personal sovereignty is regarded as an all or nothing concept. For Feinberg, one is entitled to "absolute control of whatever is in one's domain, however trivial it may be" because, he argues, this is the only conception of personal autonomy that is consistent with a "true presumption in favour of liberty" in which the state is viewed as a creation of the people and the people as free to live their own lives until a convincing argument to the contrary can be made (Feinberg 1989: 55). Accordingly, self-regarding decisions, that is, those which primarily and directly affect only the interests of the decision-maker, fall within the personal domain and hence outside the scope of legitimate state interference.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

But Feinberg’s strategy does not rule out all paternalistic measures, for it is only an individual’s truly voluntary choices that are inviolable. The concept of voluntary choice is therefore critical to Feinberg’s account and he therefore permits “soft paternalistic” strategies. Soft paternalism allows the state to intervene paternalistically when the individual’s conduct is *non-voluntary* so that there is a sense in which the individual’s choice is not really his own. Thus paternalistic interventions may be justified to ensure that the choice in question was genuinely voluntary, either because it is substantially non-voluntary, or when temporary intervention is necessary to establish whether it is voluntary or not (Feinberg 1989, chapter 20). Thus, in situations where there is a strong presumption that no normal person would voluntarily choose or consent to the kind of conduct in question, that person may be restrained until the voluntary character of his choice can be established. To use Mill’s famous example, if either a public officer or any one else saw a person attempting to cross a bridge which had been ascertained to be unsafe, and there was no time to warn him of his danger, they might seize him and turn him back, without any real infringement of his liberty. As Mill puts it, “liberty consists in doing what one desires, and he does not desire to fall into the river” (Mill 1859: 88). But if the choice in question is genuinely voluntary, then interference with that choice (at least in the form of coercive laws backed by sanctions) cannot be justified. Hence the smoker who has full awareness of its health risks cannot be interfered with. For Feinberg, the way for the state to assure itself that such practices are truly voluntarily is to continually confront smokers with the ugly medical facts so that they cannot escape the knowledge of exactly what the risks to health are (Feinberg 1989: 116).

ii. *Autonomy and “pervasive” or critical life choices*

Feinberg’s sovereign right conception of personal autonomy contrasts vividly with the conception of autonomy advocated by Gerald Dworkin, who observes that there is no single conception of autonomy, but rather one concept and many conceptions (Dworkin 1988). He favours a relatively thin conception, defining autonomy as the “second order capacity of persons to reflect critically upon their first order preferences, desires, wishes and so forth, and the capacity to accept or attempt to change these in light of higher order preferences and values” (Dworkin 1988:20). For Dworkin, although autonomy is important, so also are other capacities, such as the capacity for sympathetic identification with others, the capacity to reason prudentially, or the virtue of integrity. He thus implicitly rejects Feinberg’s sovereign right conception of autonomy, drawing the boundaries of the domain of personal sovereignty more narrowly, confining its sphere of protection to critical life choices that determine one’s lot in life. Thus (albeit with some reluctance) Dworkin accepts that we are justified in imposing coercive laws that require motorcyclists to wear helmets and sailors to take along life preservers

because, in the end, it minimises the risk of harm to them at the cost of a trivial interference with freedom (Dworkin 1988: 127).²⁸ In other words, trivial interferences with self-regarding choices may be legitimate if they are outweighed by considerations of harm diminution which the interference is expected to confer upon the individual whose freedom is restricted. Feinberg is scathing of such “balancing strategies” due to the unavoidable difficulties associated with applying such a restricted concept of personal sovereignty in non-arbitrary ways. For Feinberg, a trivial interference with sovereignty is like a minor invasion of virginity: it is respected in its entirety or not at all. On his account, personal sovereignty is not simply another value to be weighted in a cost-benefit comparison (Feinberg 1989: 20).

iii. *A third way: a sovereign right conception subject to a de minimis exception?*

Between the positions taken by Feinberg and Dworkin on whether coercive interferences with individual autonomy may be justified for paternalistic reasons are various intermediate positions, such as that of Simester and von Hirsh. In analysing paternalistic laws backed by criminal sanctions, they explicitly endorse Feinberg’s sovereign right conception of autonomy, rejecting richer conceptions of “deep autonomy” which rest on an understanding of autonomous individual action only when it accords with an individual’s coherent set of goals as part of a life conception plan of his own (Simester and von Hirsh 2011: 153). For them, the self-regarding choices of *all* competent individuals are worthy of respect and cannot therefore be interfered with by the state, particularly in the form of coercive laws backed by criminal sanctions. Yet they are nevertheless anxious to allow some kind of balancing of interests, especially for minor intrusions that do not unduly restrict meaningful options yet have potential substantial effects on victim’s lives and long term goals, particularly given that humans are fallible and may be tempted in moments of stress to take actions that lead to drastic and irreversible consequences (*ibid*). They reconcile these two positions by advocating a *de minimis* exception to the prohibition on paternalistic criminal prohibitions. Within Anglo-American legal systems, they observe that self-injurious conduct is regulated through criminal laws (seat belt laws, helmet laws), but the sanctions are limited to modest fines, imposed primarily to give some deterrent effect to the rule. Accordingly, in light of the low levels of sanction and the absence of any significantly censuring response that is typically associated with criminal sanctions, they argue that considerations of degree are relevant. Hence if a coercive intervention requires only minor precautions and is not unduly onerous or restrictive of meaningful options, while the potential effects on victim’s lives and long term goals are

²⁸ Dworkin’s view of autonomy therefore chimes with that of Joseph Raz who argues that only “pervasive” choices that are essential for autonomy, such decisions affecting one’s ability to choose one’s own career or feel part of a political community (Raz 1986: 409)

substantial, then such interventions can be justified. Moreover, they argue that the risks these measures are concerned with are not (typically) the subject of considered choices by individuals, but represent instead casual “miscalculations” that might be legitimately addressed through intervention (Simester and von Hirsh 2011: Part IV).

(c) Paternalistic choice architecture?

One striking features of philosophical reflection concerning paternalism is that it has largely focused on legal measures entailing coercive threats that are grounded in the state’s legitimate authority to deploy force in appropriate circumstances, or what regulatory scholars often refer to as “command and control” techniques for influencing human behaviour. The use of other instruments or techniques of influence underpinned by paternalistic motives, in ways that do not entail the use of coercive threats has, at least until the publication of *Nudge*, been relatively neglected. One of Thaler and Sunstein’s most important contributions has been to reinvigorate and widen scholarly debates about paternalism and its legitimacy. The scholarly approaches sketched here vividly demonstrate considerable differences in the extent to which scholars regard paternalistic state measures as justified. For some, in cases of conflict between individual autonomy and welfare, the former must triumph so that coercive choice architecture that directly interferes with an individual’s self-regarding choices is illegitimate. But even on these accounts, intentional choice architecture of a non-coercive kind (such as nudges and inducements) might be acceptable if they help ensure that an individual’s decisions are voluntary: i.e. based on a full and informed basis, in which the individual’s judgment is not clouded by diverting emotions or impulses or subject to other influences that might distort the individual’s processes of self-reflection and deliberation. Those scholars who are more willing to tolerate coercive paternalistic measures might also be more willing to regard non-coercive forms of choice architecture as legitimate, particularly if the potential to enhance the welfare of those whose behaviour the choice architect seeks to target may be significantly enhanced as a result.

5. Conclusion

This paper has explored the extent to which three different forms of choice architecture - coercion, inducements and nudge, are consistent with the fundamental values and premises upon which liberal democratic states rest. Although a variety of theories parade under the banner of liberalism, they are united in the special status which they accord to individual liberty and freedom. By examining the underlying mechanisms through which each form of choice architecture is intended to work, I have highlighted different ways in which individual freedom

can be understood: as freedom of choice, as freedom from interference (negative liberty) or positive freedom (autonomy). Each form of choice architecture implicates individual freedom in different ways that may undermine the true voluntariness of the targeted individual's decision. The individual who undertakes a particular action to avoid the unwelcome consequences threatened by the coercive choice does so on a rational basis but in circumstances that render her decision defective and therefore involuntary. In a rather different vein, the individual whose is nudged into taking particular action unreflectively in response to choice architecture which formally preserves her freedom of choice but seeks to harness the cognitive deficiencies associated with System 1 thinking may act in a non-voluntary manner, so that her actions lack voluntariness due to an operative form of deception, arguably interfering with her positive freedom while formally preserving her freedom of choice. In contrast, individuals who succumb to attractive inducements to act in particular ways do so voluntarily unless the inducement distorts the individual's evaluation of the risks and benefits of the conduct in question, so that the resulting decision may be regarded as not truly autonomous. Accordingly, each form of choice architecture raises ethical concerns from the perspective of liberal theory, although the conditions under which those concerns might be overcome is a matter of considerable contestation, as my discussion of both the harm principle and paternalism attests.

While a liberal perspective is a valuable lens for evaluating the legitimacy of choice architecture, it suffers from several weaknesses which are worth noting and warrant further exploration. In particular, at the heart of the Western liberal paradigm is an idealised conception of the self as an autonomous, rational, subject possessed of abstract liberty rights that are presumed capable of exercise and the capacity for rational deliberation, most notably the capacity rationally to scrutinise one's attachments, values and commitments detached from each of those aspects themselves. It is this model of the autonomous person that communitarians and various theorists of identity (including feminist theorists) have criticised, claiming that the notion of an unencumbered, abstracted self upon which liberal theory rests fails to reflect our lived experience and posits both an unrealistic and impoverished view of the self-society relation (Christman 2001). This insight may help explain why identifying what counts as an "interference" with the liberty or autonomy of agents is often tricky (Dworkin 2014). Although I have sought to articulate some broad guidelines for determining liberal limits on the use of choice architecture, applying them to particular real-world contexts is unlikely to be straightforward. As critical sociologists have emphasised, the abstract, liberal self is far removed from the embodied, socially situated individual whose behaviours are shaped and take place within everyday interactions with others and with their broader cultural and social environment, arguing that embodied perception and performance belong at the centre of the self-society

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

relation (Bourdieu 1984). On this understanding, the intentional shaping of choice architecture may have far more troubling implications for projects of individual self-creation than the western liberal tradition has hitherto recognised (Cohen 2012).

This paper has focused on the use of choice architecture by the state as an instrument for shaping the behaviour and decision-making of citizens. But, while the state may have a monopoly on the use of the law’s coercive power, it enjoys no such monopoly on the use of inducements or nudge techniques, as advertisers and marketing experts are well aware. Traditional advertising via the mass media can be understood as a rather blunt form of nudge: shaping the information environment of consumers by drawing attention to the products and their claimed benefits through a general advertising campaign. But the advent of networked, digital communications technologies and so-called “Big Data” techniques provide commercial providers with a much more powerful, real-time, continuously updated technique for shaping the information environment in highly personalised ways via networked information technologies that have become both ubiquitous and indispensable in the daily lives of those in advanced capitalist societies (Zuboff 2015). It is the use of nudge techniques in this networked, dynamic and personalised form that arguably provides a much more powerful and potentially worrying mode of influence (See Ford 2000; Yeung 2015a). For example, in 2014 the social media giant Facebook manipulated the newsfeeds of 700,000 users without their knowledge or consent, in a deliberate attempt to influence their emotions, provoking a storm of protest when it was subsequently made public (Arthur 2014). It is the use of choice architecture by powerful information intermediaries within networked digital environments that warrants urgent and critical scrutiny – but which traditional liberal theory may have relatively little to offer by way of guidance and constraint (Cohen 2012). Hence a significant challenge facing liberal theorists is to provide a conceptually rigorous and relevant framework for evaluating the legitimacy of the kinds of techniques for influencing individual decision-making in a globalised, networked information society which will provide concrete guidepost for shaping and constraining the use of such techniques in ways that remain true to the liberal commitment to individual self-development and a flourishing democratic community.

References

Anderson, Scott. 2015. Coercion. In *The Stanford Encyclopedia of Philosophy*, Summer 2015 Edition, edited by Edward N. Zalta (ed.), <http://plato.stanford.edu/archives/sum2015/entries/coercion/> (accessed November 27, 2015).

Anderson, Scott. 2010. "The Enforcement Approach to Coercion," *Journal of Ethics & Social Philosophy* 5: 1-31.

Arnold, N. Scott. 2009. *Imposing Values*. New York: Oxford Univ. Press.

- Ashworth, Andrew and Lucia Zedner. 2014. *The Preventive State and Its Limits*. Oxford: Oxford Univ. Press.
- Baldwin, Robert, Martin Cave, and Martin Lodge. 2012. *Understanding Regulation*, 2nd ed. New York: Oxford Univ. Press.
- Berlin, Isaiah. 1998. *Two Concepts of Liberty*. In *The Proper Study of Mankind*, edited by H. Hardy and R. Hausheer, Pimlico: London.
- Black, Julia. 2008. "Constructing and Contesting Legitimacy and Accountability in Polycentric Regulatory Regimes," *Regulation & Governance* 2: 137-64.
- Bleckmann, A, and M Bothe. 1986. "General Report on the Theory of Limitations on Human Rights." In *The Limitation of Human Rights in Comparative Constitutional Law*, edited by A de Mestral, S Birks, M. Bothe, I. Cotler, D. Klinck and A. Morel. Quebec: Les Editions Yvon Blais Inc.
- Bourdieu, Pierre. 1984. *Distinction: A Social Critique of the Judgment of Taste*. Translated by Richard Nice. Cambridge, MA: Harvard Univ. Press.
- Bovens, Luc. 2008. "The Ethics of Nudge". In *Preference Change: Approaches from Philosophy, Economics and Psychology*, edited by T. Grune-Yanoff and S. Hansso. Springer: Dordrecht, Heidelberg.
- Breyer, Stephen. 1982. "Alternatives to Classical Regulation." In *Regulation and Its Reform*, edited by Stephen Breyer. Cambridge, MA: Harvard University Press.
- Brownsword, Roger. 2005. "Cloning, Zoning and the Harm Principle." In *First Do No Harm*, edited by Sheila McLean. Aldershot: Ashgate.
- Buckley, Francis H. 2009. *Fair Governance - Paternalism and Perfectionism*. Oxford: Oxford Univ. Press.
- Cane, Peter. 2002. *Responsibility in Law and Morality*. Hart Publishing: Oxford and Oregon Portland.
- Christman, John. 2001. "Liberalism, Autonomy and Self-Transformation." *Social Theory and Practice* 27: 185-206.
- Christman, John. 2005. "Saving Positive Freedom," *Political Theory* 33:79-88.
- Christman, John. 2015. "Autonomy in Moral and Political Philosophy." In *Stanford Encyclopedia of Philosophy*, Spring 2015 Edition, edited by Edward N. Zalta at <http://plato.stanford.edu/entries/autonomy-moral/> (accessed November 27, 2015).
- Cohen, Julie E. 2012. *Configuring the Networked Self*. New Haven: Yale Univ. Press.
- Cooter, Robert. 1984. "Prices and Sanctions," *Columbia Law Review* 84: 1523-1560.
- Dayan, E. and Bar-Hillel, M. 2011. "Nudge to Nobesity II: Menu Positions Influence Food Orders," *Judgment and Decision Making* 6: 333-42.
- Duff, R. A., and S. E. Marshall. 2000. "Benefits, Burdens and Responsibilities: Some Ethical Dimensions of Situational Crime Prevention." In *Ethical and Social Perspectives on*

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Situational Crime Prevention, edited by Andrew von Hirsch, David Garland and Alison Wakefield. Portland, Oregon: Hart Publishing.

Dworkin, Gerald. 1988. *The Theory and Practice of Autonomy*. Cambridge: Cambridge Univ. Press.

Dworkin, Gerald. 2014. "Paternalism." In *The Stanford Encyclopedia of Philosophy*, Summer 2014 Edition, edited by Edward N. Zalta at <http://plato.stanford.edu/entries/paternalism/> (accessed November 25, 2015).

Dworkin, Ronald. 1985. *A Matter of Principle*. Cambridge, MA: Harvard Univ. Press.

Feinberg, Joel. 1989. *The Moral Limits of the Criminal Law Volume 3: Harm to Self*. New York: Oxford Univ. Press.

Ford, Richard T. 2000. "Save the Robots: Cyber Profiling and Your So-Called Life," *Stanford Law Review* 52: 1573-84.

Galle, Brian. 2012. "The Tragedy of the Carrots: Economics and Politics in the Choice of Price Instruments," *Stanford Law Review* 64: 797-849.

Grant, Ruth. 2011. *Strings Attached: Untangling the Ethics of Incentives*. New Haven: Princeton Univ. Press.

Haile, Andrew J. 2009-2010. "Sin Taxes: When the State Becomes the Sinner," *Temple Law Review* 82: 1041-69.

Hansen, P. G. and A. M. Jespersen. 2013. "Nudge and the Manipulation of Choice - A Framework for the Responsible Use of the Nudge Approach to Behaviour Change in Public Policy," *European Journal of Risk Regulation* 1:3-26.

Harlow, Carol and Richard Rawlings. 2009. *Law and Administration*. 3rd ed. London: Cambridge Univ Press.

Hausman, Daniel M, and Brynn Welch. 2010. "Debate: To Nudge or Not to Nudge," *Journal of Political Philosophy* 18: 123-36.

Hodkinson, Kate. 2013. "The Need to Know - Therapeutic Privilege: A Way Forward," *Health Care Analysis* 21: 115-129.

Husak, Douglas N. 2005. "Legal Paternalism." In *The Oxford Handbook of Practical Ethics*, edited by Hugh La Follette. Oxford: Oxford Univ. Press.

Jones, Peter. 1994. "Rights". In *Issues in Political Theory*, edited by Peter Jones and Albert Weale. London: Macmillan.

Kahneman, Daniel. 2013. *Thinking, Fast and Slow*. New York: Farrer, Strauss and Giroux.

Kleinig, John. 1978. "Human Rights, Legal Rights and Social Change." In *Human Rights - Ideas and Ideologies*, edited by Eugene Kamenka and Alice Erh-Soon Tah. London: Edward Arnold.

Kleinig, J., *Paternalism*. 1983, Manchester: Manchester Univ. Press.

Lamond, Grant. 2000. "The Coerciveness of Law," *Oxford J Legal Studies* 20: 39-62.

- Largent, Emily, Christine Grady, Franklin G. Miller and Alan Wertheimer. (2013) "Misconceptions About Coercion and Undue Influence: Reflections on the Views of IRB Members," *Bioethics* 27: 500-507.
- Linder, S.H. and B.G. Peters. 1989. "Instruments of Government: Perceptions and Contexts," *Journal of Public Policy* 9: 35-58.
- Lodge, Martin and Kai Wegrich. 2012. *Managing Regulation*. London: Palgrave Macmillan.
- Metrix. 2014. *The Psychology of Nudge Marketing*.
<http://www.metrixmarketing.co.uk/blog/nudge-marketing-psychology/> (accessed November 26, 2015)
- Mill, J.S. 1859. *On Liberty*. 2nd ed. London: John W Parker & Son.
- Mols, Frank, S Alexander Haslam, Molanda Jetten, and Niklas K Steffens. 2015. "Why a Nudge Is Not Enough: A Social Identity Critique of Governance by Stealth," *European Journal of Political Research* 54: 81-98.
- Morgan, Bronwen and Karen Yeung. 2007. *An Introduction to Law and Regulation*. Cambridge: Cambridge Univ. Press.
- Nozick, R., 1969. "Coercion." In *Philosophy, Science, and Method: Essays in Honor of Ernest Nagel*, edited by S. Morgenbesser, P. Suppes, and M. White. St Martin's Press: New York.
- Nuffield Council on Bioethics. 2007. *Public Health*, London.
- Nys, Thomas R. V. 2008. "Paternalism in Public Health Care," *Public Health Ethics* 1: 64-72.
- Ogus, Anthony. 1994. *Regulation - Legal Form and Economic Theory*. Oxford: Oxford Univ Press.
- Organ Donation Taskforce. 2008. *The Potential Impact of an Opt Out System for Organ Donation in the UK*, London.
- Parker, Christine. 2016. "The Food Label as Contested Nudging Space," *Law & Policy*, this volume.
- Pigou, A. C. 1932. *The Economics of Welfare*. 4th ed. London: Macmillan & Co.
- Pope, Thaddeus Mason. 2004. "Counting the Dragon's Teeth and Claws: The Definition of Hard Paternalism," *Georgia State University Law Review* 20: 659-722.
- Quigley, Muireann. 2013. "Nudging for Health: On Public Policy and Designing Choice Architecture," *Medical Law Review* 21: 588-621.
- Raz, Joseph. 1986. *The Morality of Freedom*. Oxford: Clarendon Press.
- Saghai, Yashar. 2013. "Salvaging the Concept of Nudge," *Journal of Medical Ethics* 39: 487-93.
- Sandel, Michael J. 2012. *What Money Can't Buy - The Moral Limits of Markets*. London: Penguin Books.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Sandmo, Agnar. 2008. "Public Goods." In *The New Palgrave Dictionary of Economics*, edited by Steven N Durlauf and Lawrence E Blum. Basingstoke: Palgrave Macmillan.

Scanlon, T.M. 1988. "The Significance of Choice." In *The Tanner Lectures on Human Values*, vol. 7, edited by S. McMurry. Salt Lake City, Utah: Univ. of Utah Press.

Simester, A.P. and Andrew von Hirsch. 2011. *Crimes, Harms, and Wrongs*. Oxford: Hart Publishing.

Steiner, Henry J., and Philip Alston. 2000. *International Human Rights in Context - Law, Politics, Morals*. 2nd ed. Oxford: Oxford Univ. Press.

Sunstein, Cass. 2013. "The Storrs Lectures: Behavioural Economics and Paternalism," *Yale Law Journal* 122: 1826-1862.

Thaler, Richard and Cass Sunstein. 2008. *Nudge*. London: Penguin Books.

Waldron, Jeremy. 1987. "Theoretical Foundations of Liberalism," *The Philosophical Quarterly* 37 (147): 127-50.

Waldron, Jeremy. 2014. "It's All for Your Own Good." *New York Review of Books*, October 9.

Wendell Holmes Jr, Oliver. 2004. *The Common Law*. New York: Barnes & Noble Publishing.

Wertheimer, Alan. 1988. *Coercion*. Princeton: Princeton Univ. Press.

Wertheimer, Alan. 1993. "A Philosophical Examination of Coercion for Mental Health Issues," *Behavioural Sciences and the Law* 11: 239-258.

White, Mark D. 2010. "Behavioural Law and Economics: The Assault on the Consent, Will and Dignity." In *Essays on Philosophy, Politics & Economics: Integration and Common Research Projects*, edited by Gerald Gaus, Christi Favor and Julian Lamont. Stanford, California: Stanford Univ. Press.

White, Mark D. 2013. *The Manipulation of Choice: Ethics and Libertarian Paternalism*. New York: Palgrave Macmillan.

Wilkinson, T. M. 2013. "Thinking Harder About Nudges," *Journal of Medical Ethics* 39: 486.

Yeung, Karen. 2008. "Towards an Understanding of Regulation by Design." In *Regulating Technology*, edited by Roger Brownsword and Karen Yeung. Portland, Oregon: Hart Publishing.

Yeung, Karen. 2012. "Nudge as Fudge," *Modern Law Review* 75: 122-48.

Yeung, Karen. 2015. "Design for Regulation." In *Handbook of Ethics, Values and Technological Design*, edited by Jeroen van den Hoven, Ibo van de Poel and Pieter E Vermaas. Dordrecht: Springer.

Yeung, Karen. 2015a. "Nudge on Steroids: Big Data as a Mode of Governance by Design." Paper presented at 'Algorithms and Accountability' conference, NYU Information Law Institute, New York, February 28.

Zuboff, Shoshana. 2015. "Big Other: Surveillance Capitalism and the Prospects of an Informal Civilization," *Journal of Information Technology* 30: 75-89.

(Choice architecture as control Revised for publication.docx)

For Review Only